

Review Draft



United States Department of Agriculture

**Service Center
Data Management Deliverables
For Application Development Projects**

May 9, 1999

Foreword

This document describes the data management deliverables that the Service Center Data Team expects to receive from application development projects chartered under the Business Process Reengineering (BPR) initiative.

The Service Center Data Team (Data Team) developed the material contained herein. This team is composed of representatives from NRCS, FSA and Rural Development.

Sponsorship and direction for the Data Team comes from the Executive Director of the National Food and Agriculture Council (FAC) of the U.S. Department of Agriculture (USDA) and the partner Service Center agency's Chief Information Officers (CIO's).

Data Management Deliverables for Application Development Projects

Introduction

Data Management (the function of managing data as an asset) is a shared function between the Business-Area sponsoring an application and the application developers. The Business-Area or Mission-Area has ultimate responsibility for a system, and the data it produces, since an application is developed to support a particular business function.

However, application developers have a responsibility to:

- establish the data context for the system, i.e. show how a system and its data fits into the overall enterprise,
- advise and educate business areas on the procedures to define and manage their data,
- ensure that the validity of data is maintained,
- ensure that a concise and unambiguous definition is created for each piece of data
- take reasonable steps to protect data from being used in ways that invalidate its meaning,
- ensure that the system integrates with the enterprise data model, and
- ensure that system designs promote sharing and reuse of data among systems, business disciplines, and by outside customers.

Functions and Responsibilities

This document primarily describes data management activities and documentation related to Information Systems. The data produced by a system is to be treated as an agency asset. It is an asset usually of considerable value, both monetarily and for its ability to enhance and support agency functions.

Establishing and maintaining the value of this asset starts with the initial establishment of system requirements and pervades the system development cycle. Data Management does not end when a system is deployed. It is usually just the beginning. The Business Area and the Application Developers should develop a robust plan for the use, management, and protection of the data produced by the system. Establishing Data Stewards to care for the data and to ensure the validity of data in the system is essential.

It is important for Application Developers to know in advance the amount of effort a project must budget for data management deliverables. The workload to produce these deliverables has in the past often not been calculated into project budgets.

Software Development Lifecycle

There are many varieties of life-cycle methodologies for application development projects. In this document, a generic software lifecycle is utilized which encompasses the general phases found in most projects. The phases are Initial Project Proposal, System Requirements Development, Physical Design of the system, Development and Coding, Software Testing, and the Deployment and Maintenance of the System. The Maintenance Phase can be quite lengthy and involve numerous changes and upgrades to the basic system over time.

Project Lifecycle & Deliverables

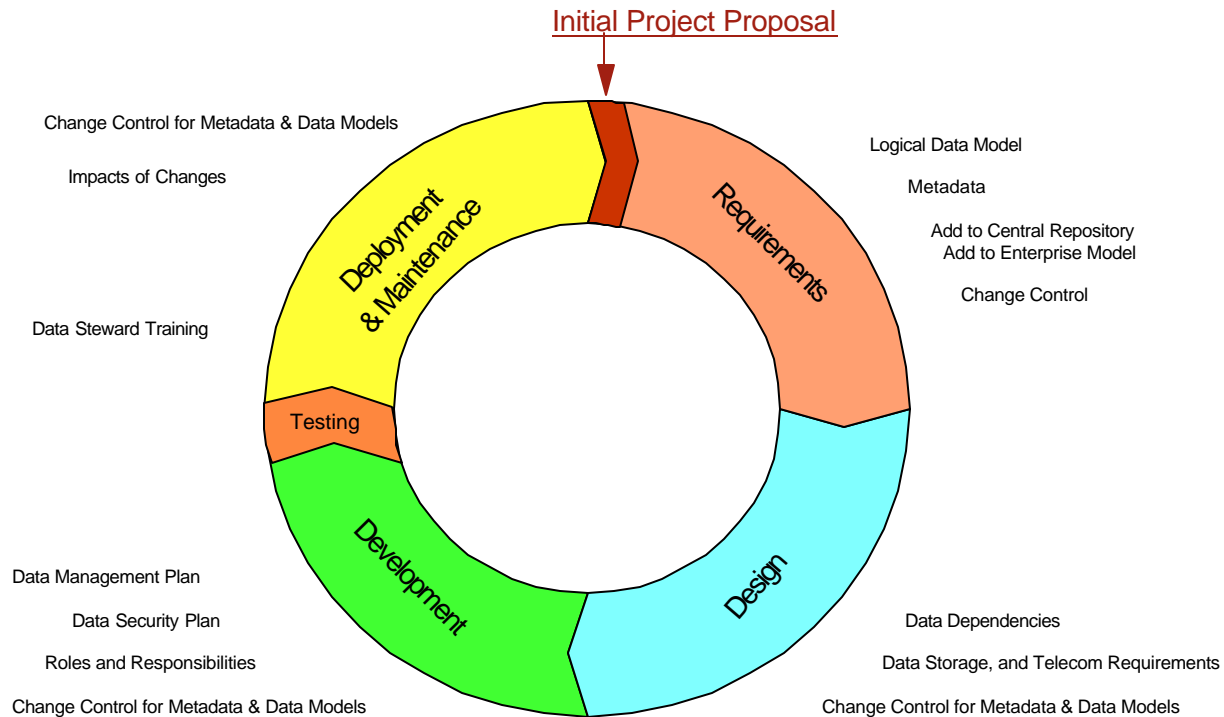


Exhibit 1 – Data Management Deliverables in each Lifecycle Phase

Data Management Deliverables

The following pages list the project deliverables for each phase of the application development lifecycle. The deliverables either require a written document or an activity such as uploading of information. These requirements have been pared down to a bare minimum needed for overall enterprise management of data resources. It is the responsibility of the Service Center Data Team to provide this overall enterprise management and to ensure coordination across the large slate of information system projects. The Data Team will be provided copies of each deliverable for review. However, each deliverable is also an essential part of system documentation, and will also be maintained with the project's documentation set.

Preparation of these deliverables is seen as a joint responsibility of the Application Developers, the Data Steward representing the sponsoring business-area, and the Service Center Data Team. There may be other data-related products needed by the application developers and data stewards to support a system in addition to those detailed in this document.

A checklist of additional data management activities that should be accomplished during each phase of project development is included.

Data Management Information for the:

Initial Project Proposal

The following chart lists Data Management deliverables to be provided in the Initial Project Proposal or for a Waiver Request by an application project team. This Proposal establishes the system's intended place in the enterprise architecture, establishes essential roles in the Information Technology and Business communities, and lays out the basic impacts of the system. Whether the system will be implemented using commercially-available packages or through a full software development project may not be known when this proposal is submitted.

	DELIVERABLES
1.	Establish project roles , including who in the business area will be the Executive Sponsor of the system, the Project Manager, and the Steward of the physical data and metadata. Include the person's position title. The Data Steward is a person from the business area.
2.	Identify special data types that the application will potentially require. These are complex data types such as photographs, scanned documents, digital signatures, spatial data, etc. that may require the purchase of new database, network (including browsers), or client/server software; or additional storage capacity.
3.	Describe special data security and data handling requirements for the system.
4.	Identify data dependencies . Will data from an existing system continue to be used or brought forward to the new system? Will other systems be dependent on data from the new system? Will the new system be dependent on data from other systems? Developers must make a good-faith effort to locate existing electronic data stores before requiring the manual origination of data.
5.	Create a high-level Business Diagram with supporting documentation indicating the main categories of data used by and created by the application, integration points with other systems, data sources external to the enterprise, and known data structures that will be shared. Indicate the types of data, such as tabular, spatial, document, binary, etc.

Data Management Review Checklist:

- Y Add a placeholder in the Enterprise Model for the application. Begin to identify existing sources of data to meet requirements outlined in the 'High-level Business Model' and the 'Data Dependencies'.
- Y Coordinate requirements for special data types and special security needs with other projects that may have similar needs.
- Y Discuss initial ideas about the technical environment envisioned for the system. The focus is on early identification of any large or highly dispersed data storage requirements so as to forecast potential load on telecommunication networks and data storage equipment; and to begin any necessary procurement processes.
- Y Identify data management resources and technical knowledge that are needed but will likely be unavailable to the project, particularly to complete the requirements and design phases. This can include skills such as data modeling, database design and tuning, warehouse design, data management, etc.

Requirements Phase

The requirements phase is where detailed business needs are identified and where system analysts convert these needs into technical system requirements. Data stores to support the application are identified, along with the individual data elements they will contain.

	DELIVERABLES
1.	Develop a logical data model . Use a CASE tool to graphically portray data elements, how they are grouped and existing relationships between elements or groups of elements. Models are not confined solely to relational databases. A flat file is a grouping of data elements, and there can be relationships between files. Hierarchical databases can also be modeled. The logical model will be incorporated into the Enterprise Data Model.
2.	<p>Create metadata to describe data elements, database tables, files, geospatial features, business areas, systems, and subsystems. The required contents for this metadata is described in the <i>Service Center Tabular Metadata Standard</i> and the <i>Service Center Geospatial Feature Metadata Standard</i>. This metadata is based on the logical data model, and will usually be captured by the project team in a CASE tool. Note that metadata for data elements is created by a project team only for new elements. If data is being used from another application, that application will have already established the required metadata. There are also established Standard Data Elements which should always be used before creating new application-specific elements. These Standard Elements can be obtained from a Data Team representative. The names of data elements will conform to the <i>Service Center Data Naming Standard</i>.</p> <p><u>For COTS-based applications:</u> When the content of the internal data stores supporting a COTS package can be accessed by a project developer, elements in that database will be included in the dictionary. Data element and data element group names established by the COTS package are not required to follow USDA naming standards.</p>
3.	Upload the Metadata and Logical Data Model to the central metadata repository.
4.	Establish Change Control over all metadata and the logical data model.

Data Management Review Checklist:

- Y Application Developers will review the Logical Data Model with a Data Team Representative. This is best done periodically throughout the requirements phase, rather than after the model has been “finalized”.
- Y Actively search for an existing electronic source for each data item before committing to collect the data.
- Y Incorporate the project’s logical model into the Enterprise Model.
- Y Application Developers will review all metadata with the Business-Area Data Steward and the Business-Area Data Administrator.
- Y Upload metadata describing data elements, data element groups, subsystems, systems, geospatial features, business areas, and stewardship to the central metadata repository. The uploaded metadata will likely be replaced by more complete information in later project phases.
- Y Identify roles and procedures for maintaining change control over metadata and the data model.

Data Management Information for the:

Design Phase

The system requirements have been fully described in logical terms, and must now be cast in physical terms to send to the programmers. Now that the system is visualized in detail, it is important to reevaluate its relationships with the outside world.

	DELIVERABLES
1.	Formalize Data Dependencies with other systems. Establish and maintain a contact list of all application systems and end-users who will have a stake in the data output from this system so that users can be notified before any modifications to the data definition, domain, or content are made. Data elements being shared from/with other systems are formally declared. Agreements should be formalized with providers of any shared data resources on which the new application is dependent.
2.	Maintain Change Control on all data items in the central metadata repository, data model, and enterprise model. Ensure that all changes to data items made during the design phase are recorded in the metadata, and that dependent users are notified in advance of any changes.
3.	Provide Detailed Requirements for data storage, telecommunication loads, and software/databases needed to support special data types. Projects should look at their expected geographical dispersion of data, data volumes, and processing loads, including the estimated average and maximum data set size and transmission frequency.

Data Management Review Checklist:

- Y Map data dependencies between systems into the Enterprise Model.
- Y Establish data structures and keys to allow for efficient movement of data between computers, for the combining of data if two offices are merged, to allow data to be extracted as a complete set, and for downloading of an office-worth of data from a national or regional server.
- Y Consider georeferencing of data for mapping and graphical representations of management information.
- Y Analyze expected data loads in conjunction with telecommunications and equipment management teams.
- Y Maintain integrity of information in the Central Metadata Repository and the Enterprise Model.
- Y Make metadata, including a data dictionary, readily available for use by developers in the next phases of the project.
- Y Employ specialists to denormalize and fine-tune the physical data model for system efficiency.
- Y Establish the complete data environment, including model and definitions, before beginning the coding phase.

Data Management Information for the:

Development Phase

This phase concentrates on: 1) documenting the system as programmed; and 2) keeping the documentation up-to-date as technical refinements are made. The Detailed Data Management Plan and the Security Plan are normal documentation produced for any project.

	DELIVERABLES
1.	Maintain Change Control on all data items in the central metadata repository, data model, and enterprise model. Ensure that all changes to data items made during the coding phase are coordinated with dependent users, and recorded in the metadata.
2.	Develop a detailed Data Management Plan . This is a detailed plan for the ongoing management of data within the business area and the application. This plan communicates roles and responsibilities to users and maintainers of data in the operational system. It sets the rules for certifying the accuracy of collected data, and for protecting data wherever it is stored. Sections of this plan may be implemented through the application's user guides and other end-user documentation. The Business-area Data Steward has primary responsibility for this deliverable.
3.	Formulate a Data Security Plan . Analyze the risk involved in the loss or corruption of data. Describe the data produced by the system, its lifespan, data retirement procedures, backup and archiving, physical data security requirements, and access rules. Include data integrity rules for preventing accidental loss of data, data corruption and the inappropriate updating or modification of data as it is transferred between offices, computers, and systems. Describe requirements for data mirroring, and for disaster recovery. Describe how record and spatial data key clashes will be handled. Describe how concurrent updates will be handled.
4.	Establish detailed Roles and Responsibilities to be performed by the system's Data Steward to maintain both the physical data and the metadata for the application. There may be additional Data Stewards responsible for portions of the data, such as a State or Local Data Steward. Also describe any additional roles to be performed, such as Database Administrator, Security Officer, Data Collector, Data Certifier, Steering Body, State IT Staff, etc.

Data Management Review Checklist:

- Y Maintain lines of communication between developers and the Service Center Data Team to ensure the integrity of the metadata and models.
- Y Perform a joint review of the Data Management Plan, the Data Security Plan, and the Roles & Responsibilities.

Data Management Information for the:

Deployment and Maintenance Phase

This phase concentrates on: 1) documenting the system as programmed; and 2) keeping the documentation up-to-date as bug-fixes and system enhancements are made.

	DELIVERABLES
1.	Maintain Change Control on all data items in the data dictionary, enterprise model and central data repository. Ensure that all changes to data items made during the deployment and maintenance phase are recorded. Maintain all data-related changes in the project CASE tool and in the central repository.

Data Management Review Checklist:

- Y Coordinate the training of Data Stewards to emphasize their role in protecting the integrity of data stores wherever they are located in the country, and for ensuring the validity of data input into the system.
- Y Fine-tune and update the Data Management Plan for the system.

Glossary

Business Area

An authorized program function or mission within an agency for which managerial responsibility has been delegated to an individual.

Business Rule

A statement that defines or constrains some aspect of the business as it is implemented in the data model (e.g. “an agency office can exist in only one location at a time”). Data-related business rules are statements, phrased in absolute terms, about data (e.g. “a telephone number must have 10 digits”), and about relationships between data (e.g. “if a phone number is entered, the phone type must also be entered”).

Change Control

Change control is an active management process stabilizes the software development environment and protects systems/users who are dependent on particular data from being adversely impacted by changes to the definition, type, availability, or content of electronic data. The basic prerequisites for providing change control are: 1) establishment of management control over the definition of the data; 2) a means to notify users that a change is being proposed; 3) a means to gather responses from dependent users and to assess the potential business impacts of the change. Note that the users may be programmers on the project who are dependent on the format, availability, metadata, and stability of the data stores to complete their tasks.

Data

A discrete fact or value. Data is the raw material, which through its use and interpretation can provide valuable information. Data is the content of databases or data files.

Data Integrity

The state that exists when data is handled as intended and is not exposed to accidental or malicious modification, destruction, or disclosure. Data is entered using formal business rules, and accuracy is checked using automated edits and manual reviews. Also, includes the preservation of data for its intended use.

Data Management

The managerial function of taking responsibility for data and the processes that support it. This function focuses strategic planning and operational data functions on meeting program delivery goals.

Data Model

A pictorial view of data, groupings of data, and relationships between data groupings. A “logical” data model is a view that does not depend on the characteristics of the computerized system or of the physical storage. A “physical” data model typically refines the logical model by adding the constraints incumbent to the database system or physical storage method, and tuning the data model for access efficiency.

Data Steward

A business area expert who is assigned responsibility for the data content of the database. The data steward establishes business rules, defines data elements, identifies valid data values, establishes certification standards, and ensures the completeness and availability of the data.

Data Validation

Applying a set of rules, comparisons, or decisions to a data element to determine if it falls within the pre-established boundaries of values for that element.

Domain

A list of all possible valid values for a data element. The domain can alternatively be expressed as a range of numeric or alphabetic values, or as a reference to an identified standard, such as a FIPS table.

Enterprise Data Model

An overall pictorial view of the participating agency's many applications and data assets. The intent is to manage the overall data assets to achieve optimal integration, sharing, access, and utilization of technology resources and infrastructure.

Geospatial (GIS) Metadata

Metadata that describes geospatial features. Points, lines, and polygons are created in a GIS tool and have little meaning if they are not described in detail. These descriptions can include such items as a name for the feature, category (i.e. farm, field, water, wetland), creation date, crop cover, ownership, bounding coordinates, scale, etc. There is geospatial metadata, as well, that describes large sets of geospatial data. This includes metadata that describes a complete map. This type of geospatial metadata is used in particular to advertise or allow users to locate maps and geospatial datasets.

Information

A commodity derived from data through analysis or by the orderly presentation of data for human interpretation.

Metadata

Data about data. Metadata describes how, when, and by whom a particular set of data was collected, and how the data is formatted. Metadata includes attributes such as data name, length, domain of valid values, and definition. Metadata can also identify and describe a set of data or a complex data type such as a map, photograph, spatial data set, etc.

Metadata Repository

A database of information describing the characteristics (metadata) of data. Typically, the repository also stores a broad range of descriptive information, including business rules and data models that help to elaborate on the usage of data in various systems. Repositories can also store metadata for the purpose of identifying and retrieving sets of actual data. Metadata that describes a map is an example.